

10,789,351

7-19-4



**Europäisches
Patentamt**

**European
Patent Office**

**Office européen
des brevets**

Bescheinigung

Certificate

Attestation

Die angehefteten Unterla-
gen stimmen mit der
ursprünglich eingereichten
Fassung der auf dem näch-
sten Blatt bezeichneten
europäischen Patentanmel-
dung überein.

The attached documents
are exact copies of the
European patent application
described on the following
page, as originally filed.

Les documents fixés à
cette attestation sont
conformes à la version
initialement déposée de
la demande de brevet
européen spécifiée à la
page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

03425133.0

**CERTIFIED COPY OF
PRIORITY DOCUMENT**

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

THIS PAGE BLANK (USPTO)



Anmeldung Nr:
Application no.: 03425133.0
Demande no:

Anmeldetag:
Date of filing: 28.02.03
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

STMicroelectronics S.r.l.
Via C. Olivetti, 2
20041 Agrate Brianza (Milano)
ITALIE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se référer à la description.)

Voltage regulation system for a multiword programming of a low integration area
non volatile memory

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s)
revendiquée(s)

Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

G11C/

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of
filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL
PT SE SI SK TR LI

THIS PAGE BLANK (USPTO)

Titolo: Sistema di regolazione di tensione per la programmazione multiparola (multiword) in memorie non volatili a bassa occupazione di area circuitale.

DESCRIZIONE

5 Campo di applicazione

La presente invenzione fa riferimento ad un sistema di regolazione di tensione per la programmazione multiparola (multiword) in memorie non volatili, ad esempio di tipo Flash, a bassa occupazione di area circuitale.

- 10 Più in particolare, ma non esclusivamente, l'invenzione riguarda un sistema del tipo suddetto e del tipo in cui le memorie comprendono almeno una matrice di celle di memoria organizzata in righe e colonne di celle e con relativi circuiti preposti alle funzioni di indirizzamento, di decodifica, di lettura, scrittura e cancellazione del contenuto delle celle di memoria, dette celle avendo terminali di drain collegati alle colonne della
- 15 matrice e polarizzati in fase di programmazione con un predeterminato valore di tensione attraverso circuiti di program load associati a ciascuna colonna della matrice.

Arte nota

- 20 E' noto che nelle moderne memorie non volatili, ad esempio di tipo Flash EEPROM, è sempre più sentita la necessità di applicare alle celle di memoria tensioni molto "precise" in fase di scrittura.

Ciò è generalmente valido sia per scrittura di celle aventi un terminale di bulk mantenuto a massa, in sostanza con $V_{bulk}=0$, sia per la scrittura con V_{bulk} negativa.

- 25 In pratica i requisiti richiesti alla differenza di tensione $V_{drain} - V_{bulk}$ si possono rapportare ai requisiti richiesti alla differenza di tensione $V_{drain} - V_{source}$.

Con il termine "precise" s'intende in realtà il rispetto di due requisiti di ordine successivo:

- 30 In primo luogo si deve garantire che la cella flash, durante l'operazione di scrittura, lavori all'interno di una finestra di lavoro ottimale, ossia all'interno di una ben definita regione del piano V_{drain} vs L , dove L è la

lunghezza di canale della cella.

In tali condizioni si garantisce ad esempio un salto di soglia necessario per passare da una cella cancellata a una cella programmata nel tempo di programmazione richiesto.

5. Il parametro L è chiaramente variabile statisticamente all'interno di un array o di una matrice; tanto più tale parametro è centrato attorno ad un valore nominale, tanto più affidabile è il processo con cui si realizza la matrice di celle.

10. Analizziamo invece ora il secondo parametro V_{drain} . La necessità di rientrare comunque nella finestra di lavoro ottimale comporta una ridotta tolleranza su V_{drain} ($V_{\text{drain}} = V_{\text{nom}} \pm d\%$) che si ripercuote direttamente sulla tolleranza ammessa per il regolatore di tensione che deve generare la tensione di drain.

15. Nelle attuali memorie flash in realtà non si scrive mai una sola cella, bensì un numero di celle compreso tra 1 e 64.

Al variare del numero di celle da programmare, e quindi della corrente totale di carico I_{program} , il regolatore di tensione di drain deve fornire una tensione regolata che permetta a ciascuna cella di lavorare all'interno della finestra di lavoro.

20. Dunque l'obiettivo è quello di riuscire ad ottenere valori di V_{drain} quanto più possibile stabili o di poco variati al variare del numero di celle da programmare. Idealmente si vorrebbe:

25. V_{gate} =costante o rampa uguale per ogni cella e per ogni operazione di programmazione. Non ci sono grossi problemi nel rispettare questo requisito poiché il regolatore lavora sui terminali di gate delle celle Flash e quindi su un carico ad alta impedenza;

30. V_{bulk} =costante uguale per ogni cella e per ogni operazione di programmazione. Anche per questa richiesta non ci sono grossi problemi nel rispettarla poiché il regolatore lavora sul bulk delle Celle Flash e quindi su un carico ad alta impedenza (se trascuriamo la corrente di bulk di programmazione in genere molto più piccola di quella di drain);

$V_{\text{source}}=0$ per ogni cella e per ogni operazione di programmazione. Questa è una richiesta a prima vista poco rilevante, ma necessita di un buon

routing di layout e una buona polarizzazione delle linee di source.

$$V_{\text{drain}} = V_{\text{nom}} \pm d\%$$

La richiesta di precisione per le tensioni in gioco durante la programmazione è anche più esasperata in dispositivi flash di tipo avanzato, come nel caso delle memorie flash multilivello in cui l'informazione da immagazzinare in una cella non è più quella di un solo bit (per la quale è sufficiente verificare se la carica è presente/assente sulla floating gate) ma di una pluralità di bit (per cui occorre discriminare per via indiretta quanta carica è presente sulla floating gate).

Le imprecisioni sulle tensioni di programmazione contribuiscono ad aumentare le ampiezze delle distribuzioni dei bit rappresentati da ogni cella che in tal modo rendono difficile o addirittura impossibile poter discriminare tali livelli.

Come rappresentato nelle figure 1 e 2 è importante che le distribuzioni in soglia (V_t) delle celle relative ai livelli diversi siano ben distanziate fra di loro in modo tale da rendere agevole un successivo rilevamento (sensing) utilizzando come tensione discriminante (V_r) quella in corrispondenza delle linee tratteggiate.

Imprecisioni a livello di programmazione porterebbero le varie distribuzioni ad avvicinarsi e addirittura ad intersecarsi rendendo impossibile una successiva operazione di discriminazione.

La distribuzione dei livelli in modo preciso è una operazione delicata che dal lato della progettazione del circuito di memoria impone uno sforzo non indifferente.

Dal lato fisico è chiaro che si richiede una tolleranza sui parametri della cella molto più bassa.

Regolazione uno a tutti/uno ad uno

Prima di affrontare il problema regolazione, diciamo subito che in genere una progettazione è efficace se risponde almeno a tre requisiti:

Qualità;

Occupazione di area;

Tempo di realizzazione o di sintesi

In realtà questi parametri hanno dei pesi diversi, il fallimento di uno di questi può pregiudicare l'intero lavoro. Un ottimo circuito, sintetizzato in tempi troppo lunghi può diventare obsoleto prima che si realizzi e così via.

Vediamo più in dettaglio i due casi di regolazione:

- 5 Regolazione uno a tutti: è un tipo di regolazione illustrata in figura 3A.

In questo caso il regolatore alimenta direttamente tutte le celle che lo necessitano (da 1 a N_{max} , con $N_{max}=8 - 64$ o più)

- 10 Il regolatore deve quindi regolare l'uscita di una pompa di carica, normalmente infatti nelle memorie Flash le tensioni di scrittura vengono ottenute internamente con survoltori a pompa di carica e con un carico che varia da:

Nr bit	Current
1	$I_{program}$
N_{max}	$N_{max} * I_{program}$

- 15 Con una corrente dell'ordine di $100\mu A$ possiamo avere una variazione di carico da $100\mu A$ a $6.4mA$, dovendo sempre mantenere $V_{drain}=V_{nom} \pm d\%$

Regolazione uno ad uno: è un tipo di regolazione illustrato in figura 3B.

In questo caso si ha un regolatore per ogni cella da programmare.

- 20 Il regolatore deve quindi regolare l'uscita della pompa di carica con un carico che questa volta è costante:

Nr bit	Current
1	$I_{program}$

In questo caso la regolazione della tensione al valore $V_{drain}=V_{nom} \pm d\%$ è molto più semplice e precisa.

- 25 Pur vantaggioso sotto vari aspetti, questo secondo metodo ha un grave inconveniente dovuto all'occupazione di area circuitale richiesta dal regolatore, cosa che nelle moderne memorie a sempre più elevata scala di integrazione è uno dei fattori a cui prestare invece maggiore attenzione.

- 30 Il problema tecnico che sta alla base della presente invenzione è quello di escogitare un sistema di regolazione della tensione per la programmazione multiparola (multiword) in memorie Flash avente caratteristiche

strutturali e funzionali tali da richiedere una bassa occupazione di area circuitale.

Sommario dell'invenzione

5 L'idea di soluzione alla base della presente invenzione è quella di realizzare percorsi dummy di corrente associati in parallelo ai circuiti di program load ed abilitati mediante pass transistors, ad esempio con segnali di abilitazione complementari rispetto a quello di abilitazione dei program.

In questo modo possono essere evitati sbalzi della corrente di carico che affliggono tuttora le soluzioni dell'arte nota

10 Sulla base di tale idea di soluzione il problema tecnico è risolto da un sistema del tipo precedentemente indicato e definito dalla rivendicazione 1 e seguenti.

Le caratteristiche ed i vantaggi del sistema secondo l'invenzione risulteranno dalla descrizione di un esempio di realizzazione dato a titolo
15 indicativo e non limitativo con riferimento ai disegni allegati.

Breve descrizione dei disegni

- la figura 1 mostra schematicamente un diagramma che illustra la distribuzione della tensione di soglia in una convenzionale cella di memoria a un bit per cella;
- 20 - la figura 2 mostra schematicamente un diagramma che illustra la distribuzione della tensione di soglia in una cella di memoria multilivello a un due per cella;
- le figure 3A e 3B mostrano viste schematiche di una porzione di una matrice di celle di memoria in un dispositivo elettronico di memoria non-volatile con un associato decodificatore di colonna secondo l'arte
25 nota, nei casi di regolazione uno a tutti ed uno ad uno, rispettivamente;
- la figura 4 mostra una vista schematica di una porzione di una matrice di celle di memoria in un dispositivo elettronico di memoria non-volatile con un associato sistema di regolazione di tensione realizzato secondo
30 l'invenzione per la fase di programmazione;
- la figura 5 mostra a titolo comparativo andamenti di segnali interni a sistemi di regolazione realizzati secondo l'arte nota e secondo

l'invenzione.

Descrizione dettagliata

Con riferimento a tali figure, con 1 è globalmente e schematicamente indicato un sistema di regolazione della tensione per la programmazione
 5 multiparola (multiword) in memorie non volatili, ad esempio di tipo Flash, a bassa occupazione di area circuitale.

Le memorie non volatili sono dispositivi elettronici integrati su semiconduttore e dotati di matrici di celle di memoria, in particolare celle multilivello.

10 Più in particolare, per dispositivo di memoria si intende un qualunque sistema elettronico monolitico incorporante una matrice di celle 5 di memoria, organizzate in righe, dette word lines 6, e colonne, dette bit lines 4, nonché porzioni circuitali associate alla matrice di celle e preposte alle funzioni di indirizzamento, di decodifica, di lettura, scrittura e
 15 cancellazione del contenuto delle celle di memoria.

Un dispositivo di questo genere può essere ad esempio un chip di memoria integrato su semiconduttore e del tipo Flash EEPROM non volatile suddivisa in settori e cancellabile elettricamente.

20 Ciascuna cella di memoria comprende un transistor a floating gate con terminali di source S, drain D e control gate G.

Tra le porzioni circuitali associate alla matrice di celle è prevista la presenza circuiti di programmazione 2 (program load) ciascuno dei quali è associato ad una relativa colonna 4 della matrice ed alimentato da una specifica tensione di alimentazione, generata internamente al dispositivo
 25 integrato di memoria e regolata tramite un regolatore di tensione di drain Vd, per applicare alle colonne della matrice di celle opportune tensioni di polarizzazione.

Inoltre, è previsto un decodificatore di colonna 3 a cui fanno capo le colonne o bit lines 4 della matrice di celle.

30 Anche il decodificatore di colonna 3, situato a valle dei program load 2, è alimentato da specifiche tensioni generate internamente al circuito integrato di memoria mediante survoltori o pompe di carica 12 e regolate tramite relativi regolatori di tensione. Nell'esempio di figura 3 la pompa di

carica 12 è regolata dall'uscita di un regolatore 7 di tensione stabile attraverso la serie di un oscillatore controllato in tensione VCO e di un generatore di fase 8.

5 I principi della presente invenzione sono specificatamente sfruttati per effettuare una regolazione locale per settore della tensione V_d di drain in fase di programmazione.

La tensione applicata al terminale di drain D delle celle 5 è infatti un parametro che deve essere controllato con grande accuratezza durante la programmazione.

10 I principi della presente invenzione sono applicati ad un dispositivo integrato di memoria in cui le tecniche circuitali e i processi tecnologici impiegati consentono di ottenere una corrente di drain pressoché costante durante tutta l'operazione di programmazione.

15 Vantaggiosamente, secondo l'invenzione, in parallelo a ciascun circuito 2 di program load è previsto un percorso 9 di conduzione verso massa abilitato da un elemento attivo comandato.

Il percorso 9 comprende un pass transistor 10 ricevente sul proprio terminale di comando un segnale di abilitazione /PGM.

20 Vantaggiosamente, inoltre, detto segnale di abilitazione è complementare al segnale di abilitazione PGM applicato al corrispondente circuito 2 di program load.

Il percorso 9 è sostanzialmente un percorso ridondante o dummy.

25 Pertanto, l'invenzione prevede percorsi 9 dummy di corrente associati in parallelo ai circuiti di program load 2 ed abilitati mediante pass transistors 10, ad esempio con segnali di abilitazione complementari rispetto a quello di abilitazione dei program.

In questo modo il valore di corrente per il percorso 9 parallelo è facilmente impostabile, anche se il suo valore esatto non è importante per la nostra applicazione.

30 In questo caso la regolazione più semplice da realizzare in quanto il regolatore si trova sempre a funzionare in una situazione nota, ossia come se stesse programmando sempre il numero massimo di celle consentito dal dispositivo.

L'invenzione accomuna i vantaggi delle soluzioni note ma senza essere afflitta dai loro inconvenienti.

5 In particolare, il sistema dell'invenzione ha il vantaggio di una minima occupazione in area circuitale come nelle soluzioni di tipo uno a tutti, infatti richiede soltanto l'inserimento di due transistori per ciascun ramo di cella di memoria.

Il sistema secondo l'invenzione ha però anche il vantaggio della soluzione uno ad uno, vale a dire consente di effettuare una regolazione con carico praticamente costante.

10 In sostanza, questo sistema evita sbalzi della corrente di carico che affliggono le soluzioni dell'arte nota.

15 Un confronto tra i ripple della tensione di programmazione tra la soluzione dell'arte nota e la soluzione secondo l'invenzione dimostra che il ripple si mantiene contenuto e costante indipendentemente dal numero di bit da programmare come mostrato in figura 5.

RIVENDICAZIONI

1. Sistema di regolazione di tensione per la programmazione multiparola (multiword) in memorie non volatili, ad esempio di tipo Flash, a bassa occupazione di area circuitale, del tipo in cui le memorie
5 comprendono almeno una matrice di celle (5) di memoria organizzata in righe e colonne di celle e con relativi circuiti preposti alle funzioni di indirizzamento, di decodifica, di lettura, scrittura e cancellazione del contenuto delle celle di memoria, dette celle avendo terminali di drain (D) collegati alle colonne della matrice e polarizzati in fase di programmazione
10 con un predeterminato valore di tensione attraverso circuiti di program load (2) associati a ciascuna colonna della matrice, caratterizzato dal fatto di prevedere, in parallelo a ciascun circuito (2) di program load, un percorso (9) di conduzione verso massa abilitato mediante un elemento attivo (10) comandato.
- 15 2. Sistema secondo la rivendicazione 1, caratterizzato dal fatto che detto elemento attivo comandato è un pass transistor (10) ricevente sul proprio terminale di comando un segnale di abilitazione (/PGM).
3. Sistema secondo la rivendicazione 1, caratterizzato dal fatto che detto segnale di abilitazione è complementare al segnale di abilitazione
20 (PGM) applicato al corrispondente circuito (2) di program load.
4. Sistema secondo la rivendicazione 1, caratterizzato dal fatto che detto percorso (9) è sostanzialmente un percorso ridondante o dummy di corrente.

RIASSUNTO

L'invenzione riguarda un sistema di regolazione di tensione per la programmazione multiparola (multiword) in memorie non volatili, ad esempio di tipo Flash, a bassa occupazione di area circuitale, del tipo in cui le memorie comprendono almeno una matrice di celle (5) di memoria organizzata in righe e colonne di celle e con relativi circuiti preposti alle funzioni di indirizzamento, di decodifica, di lettura, scrittura e cancellazione del contenuto delle celle di memoria. Le celle di memoria hanno terminali di drain (D) collegati alle colonne della matrice e polarizzati in fase di programmazione con un predeterminato valore di tensione attraverso circuiti di program load (2) associati a ciascuna colonna della matrice; vantaggiosamente, l'invenzione prevede, in parallelo a ciascun circuito (2) di program load, un percorso (9) di conduzione verso massa abilitato mediante un elemento attivo (10) comandato.

(Fig. 4)

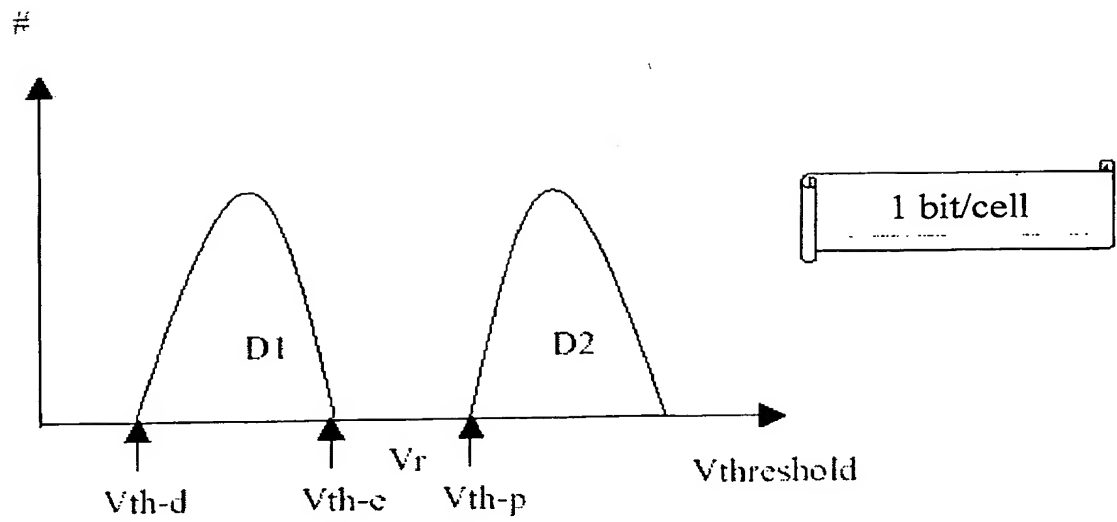


FIG. 1

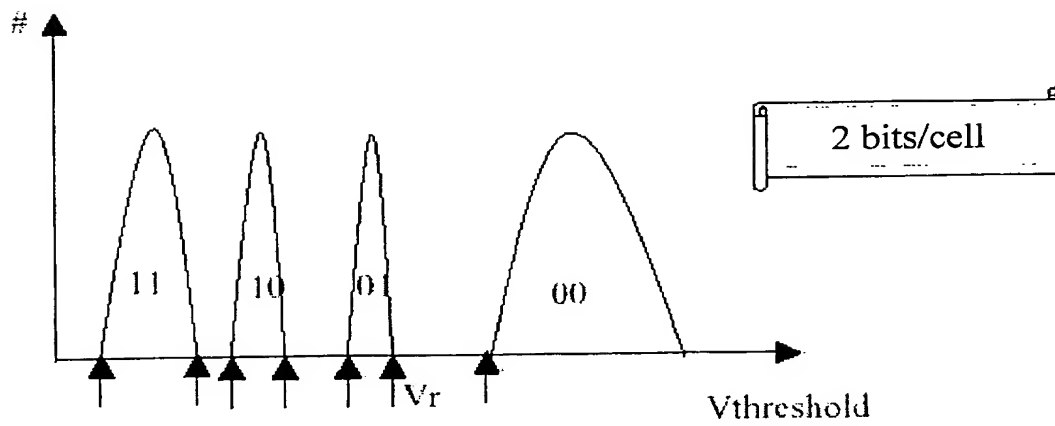


FIG. 2

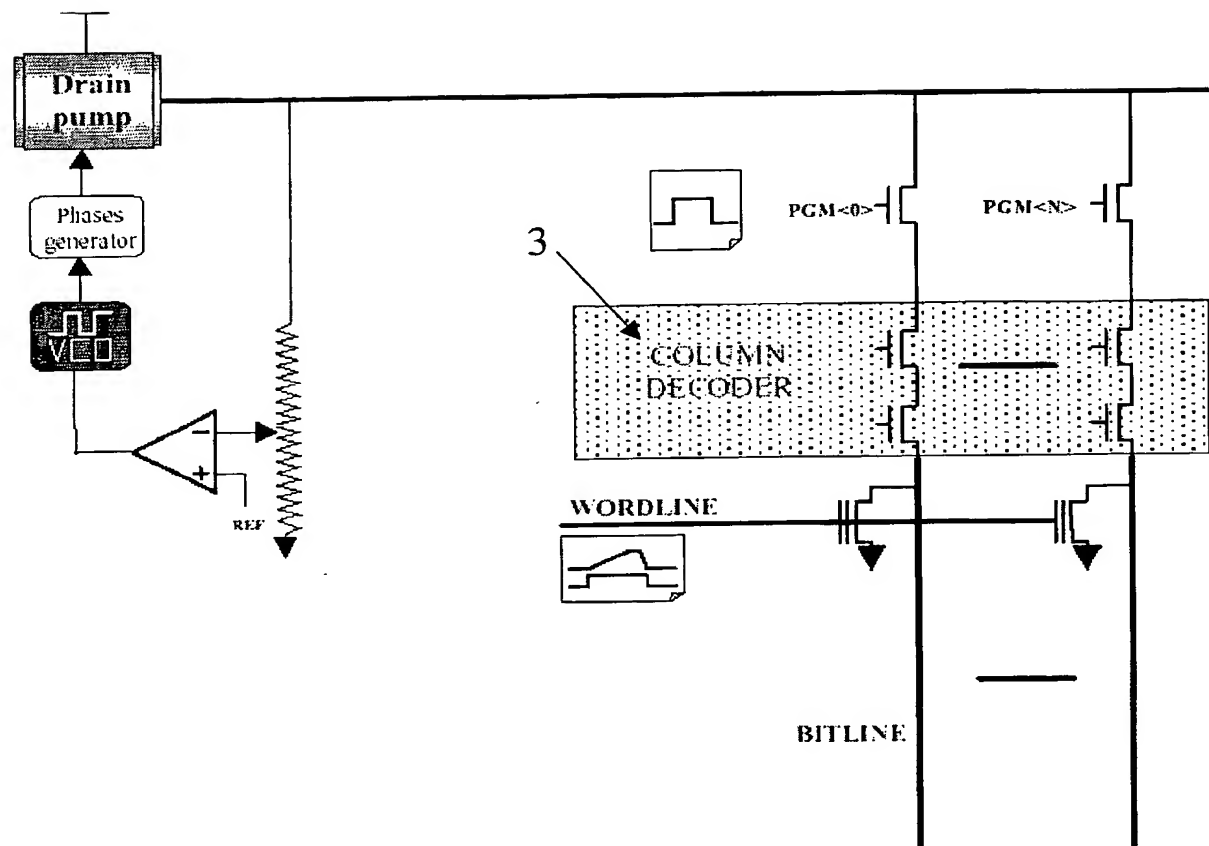


FIG. 3A

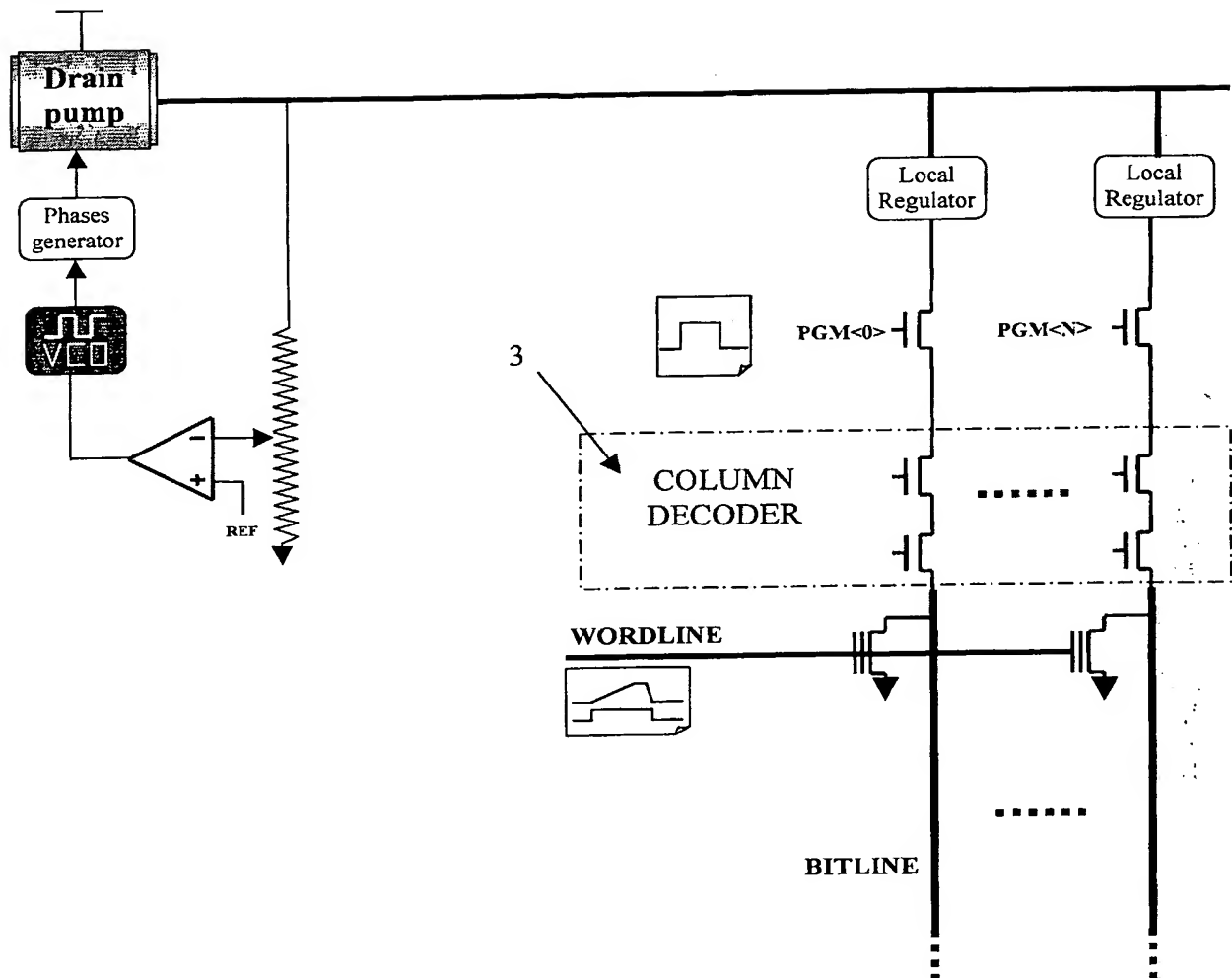


FIG. 3B

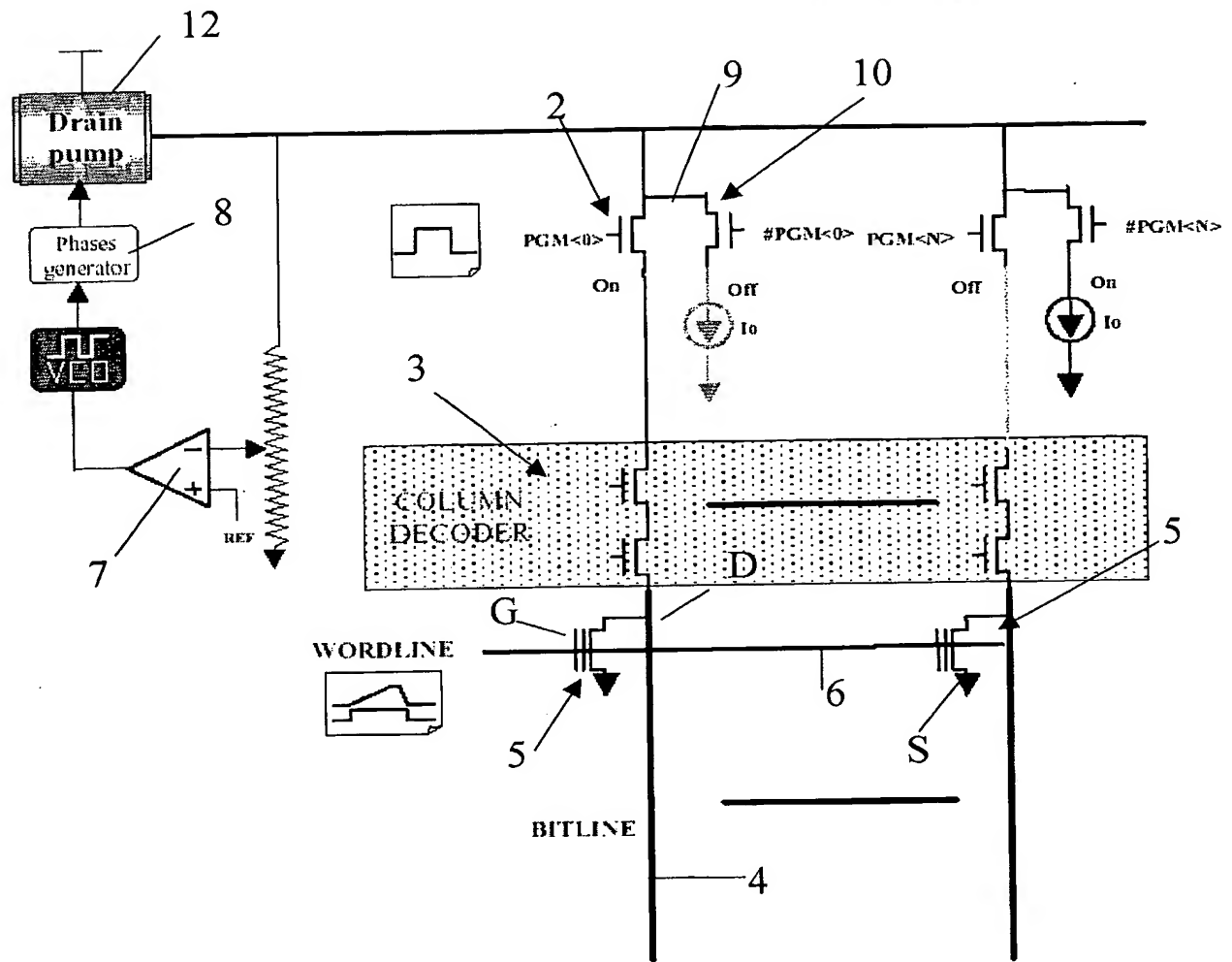


FIG. 4

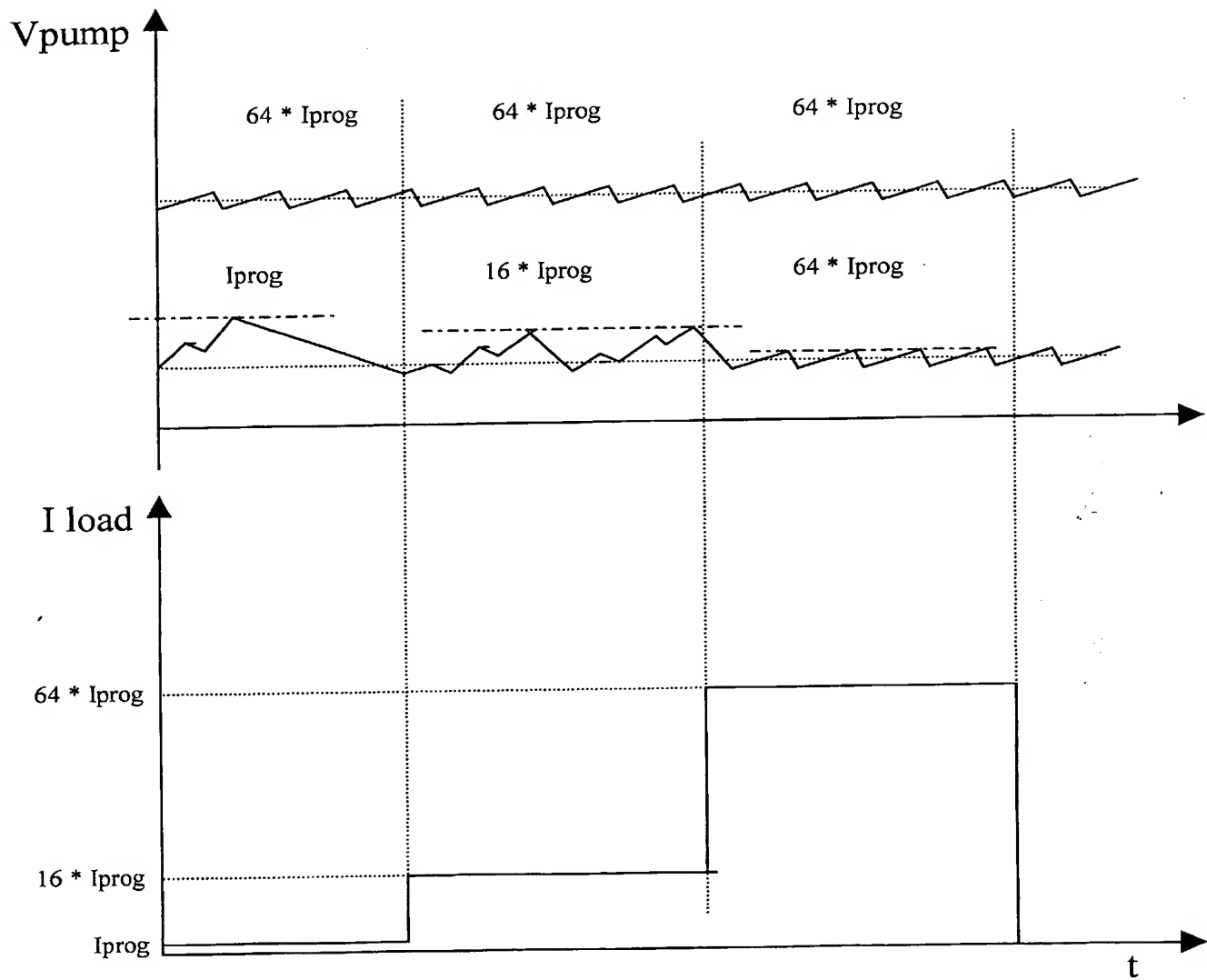


FIG. 5

THIS PAGE BLANK (USPTO)